



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:11 AM GMT

PDB ID : 2W7P
Title : STRUCTURE AND ACTIVITY OF BYPASS SYNTHESIS BY HUMAN
DNA POLYMERASE KAPPA OPPOSITE THE 7,8-DIHYDRO-8-OXOD
EOXYGUANOSINE ADDUCT
Authors : Irimia, A.; Egli, M.
Deposited on : 2008-12-23
Resolution : 3.71 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

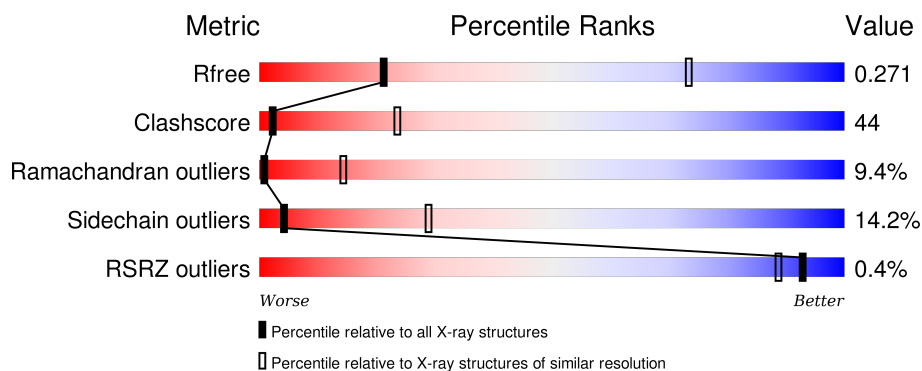
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.71 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1129 (3.94-3.50)
Clashscore	102246	1252 (3.94-3.50)
Ramachandran outliers	100387	1199 (3.94-3.50)
Sidechain outliers	100360	1197 (3.94-3.50)
RSRZ outliers	91569	1136 (3.94-3.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	508	<div> <div>30%</div> <div>44%</div> <div>10%</div> <div>15%</div> </div>
1	B	508	<div> <div>27%</div> <div>43%</div> <div>14%</div> <div>•</div> <div>15%</div> </div>
2	E	13	<div> <div>8%</div> <div>77%</div> <div>15%</div> </div>
2	P	13	<div> <div>100%</div> </div>
3	F	18	<div> <div>50%</div> <div>33%</div> <div>17%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	T	18	 A horizontal bar chart showing the quality of chain T. The bar is divided into three segments: a green segment on the left labeled '33%', a yellow segment in the middle labeled '61%', and a grey segment on the right labeled '6%'. The segments are separated by thin black lines.

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 8125 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA POLYMERASE KAPPA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	432	Total	C	N	O	S	0	0	1
			3461	2181	613	646	21			
1	B	431	Total	C	N	O	S	0	0	1
			3452	2175	611	645	21			

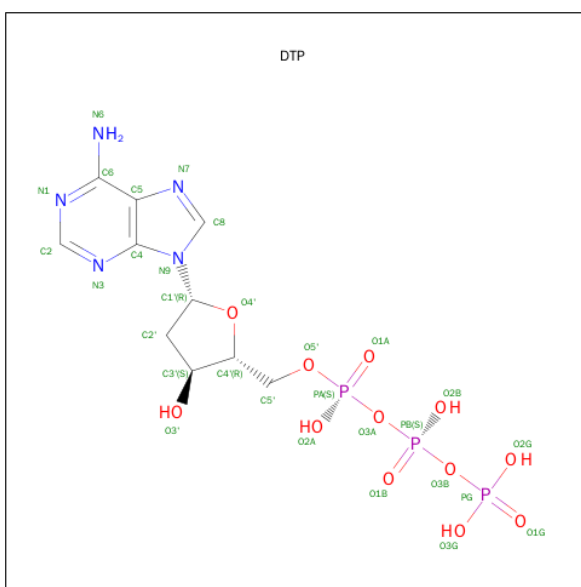
- Molecule 2 is a DNA chain called 5'-D(*GP*GP*GP*GP*GP*AP*AP*GP*GP*AP*TP*TP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	11	Total	C	N	O	P	0	0	0
			229	109	47	63	10			
2	P	13	Total	C	N	O	P	0	0	0
			273	129	57	75	12			

- Molecule 3 is a DNA chain called 5'-D(TP*CP*AP*CP*8OGP*GP*AP*AP*TP*CP*CP*TP*TP*CP*CP*CP*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	15	Total	C	N	O	P	0	0	0
			298	143	52	89	14			
3	T	17	Total	C	N	O	P	0	0	0
			336	161	58	101	16			

- Molecule 4 is 2'-DEOXYADENOSINE 5'-TRIPHOSPHATE (three-letter code: DTP) (formula: C₁₀H₁₆N₅O₁₂P₃).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			30	10	5	12	3		
4	B	1	Total	C	N	O	P	0	0
			30	10	5	12	3		

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	2	Total	Ca	0	0
			2	2		
5	A	2	Total	Ca	0	0
			2	2		

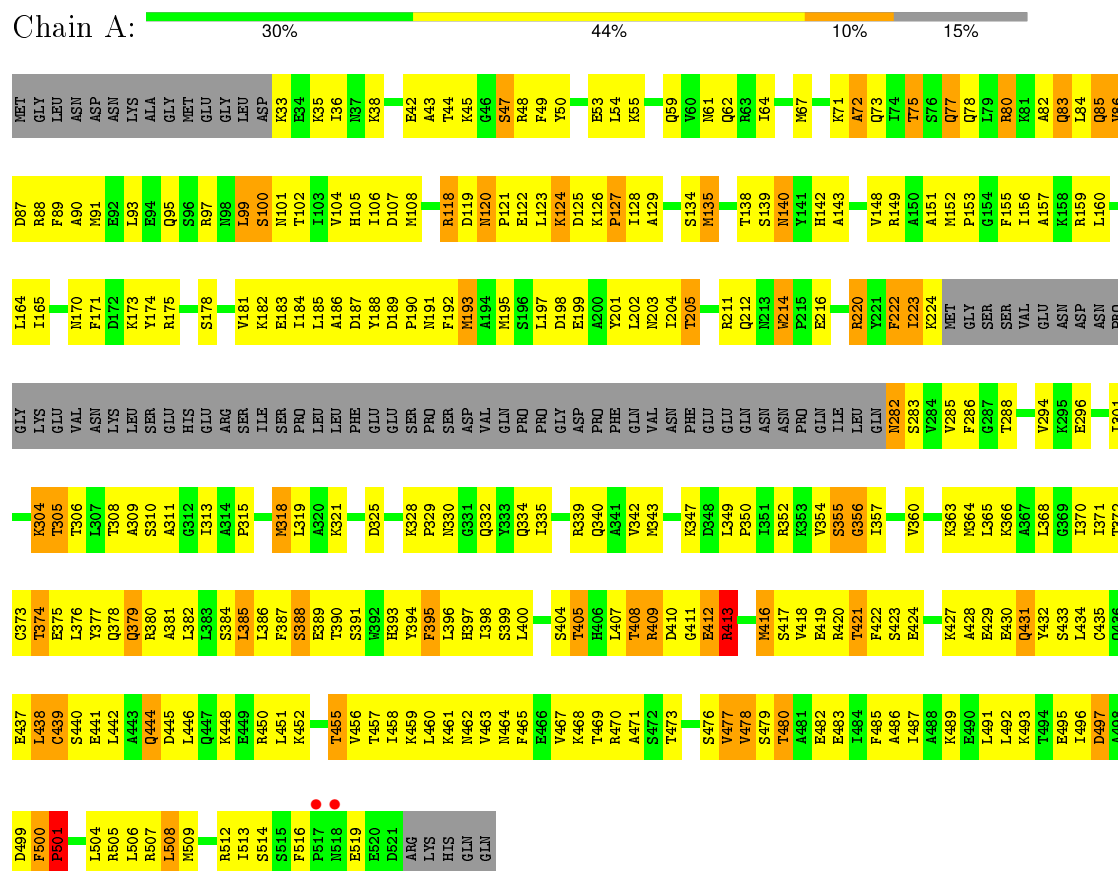
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	10	Total	O	0	0
			10	10		
6	F	1	Total	O	0	0
			1	1		
6	T	1	Total	O	0	0
			1	1		

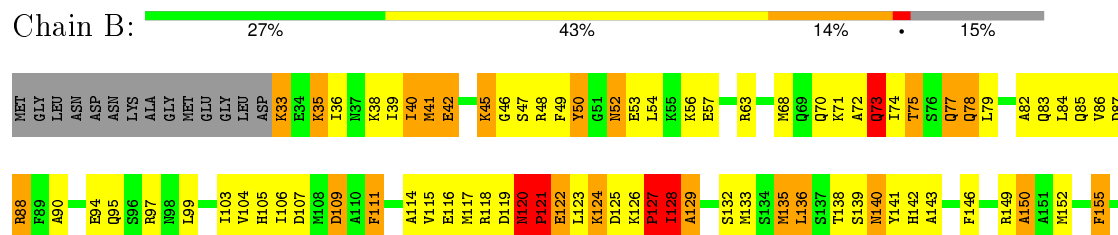
3 Residue-property plots

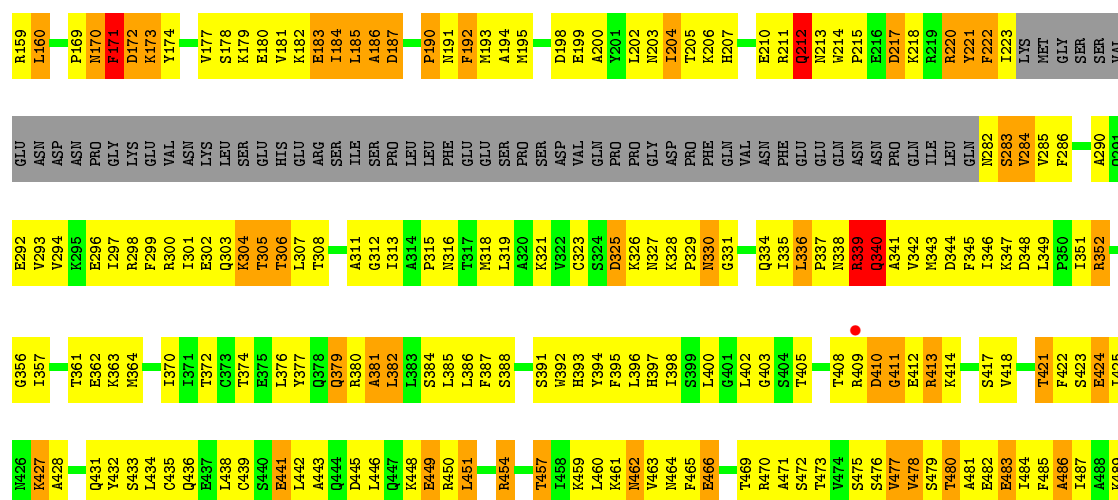
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($\text{RSRZ} > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: DNA POLYMERASE KAPPA



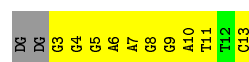
• Molecule 1: DNA POLYMERASE KAPPA





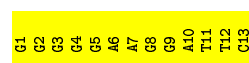
- Molecule 2: 5'-D(*GP*GP*GP*GP*GP*AP*AP*GP*GP*AP*TP*TP*C)-3'

Chain E: 8% 77% 15%



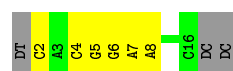
- Molecule 2: 5'-D(*GP*GP*GP*GP*GP*AP*AP*GP*GP*AP*TP*TP*C)-3'

Chain P: 100%



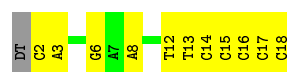
- Molecule 3: 5'-D(TP*CP*AP*CP*8OGP*GP*AP*AP*TP*CP*CP*TP* TP*CP*CP*CP*CP*C)-3'

Chain F: 50% 33% 17%



- Molecule 3: 5'-D(TP*CP*AP*CP*8OGP*GP*AP*AP*TP*CP*CP*TP* TP*CP*CP*CP*CP*C)-3'

Chain T: 33% 61% 6%



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	167.83Å 220.85Å 119.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.65 – 3.71 44.65 – 3.71	Depositor EDS
% Data completeness (in resolution range)	96.9 (44.65-3.71) 97.0 (44.65-3.71)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.62 (at 3.77Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.236 , 0.273 0.229 , 0.271	Depositor DCC
R_{free} test set	1112 reflections (5.07%)	DCC
Wilson B-factor (Å ²)	88.7	Xtriage
Anisotropy	0.723	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 41.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	0 of 23025 reflections	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	8125	wwPDB-VP
Average B, all atoms (Å ²)	95.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.56% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 8OG, CA, DTP

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z > 5$	RMSZ	$\# Z > 5$
1	A	0.48	0/3514	0.77	3/4722 (0.1%)
1	B	0.43	0/3505	0.69	0/4711
2	E	0.69	0/258	0.62	0/398
2	P	0.71	0/308	0.67	0/476
3	F	0.65	0/305	0.64	0/464
3	T	0.71	0/347	0.68	0/528
All	All	0.50	0/8237	0.72	3/11299 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1
2	P	0	1
All	All	0	2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	500	PHE	C-N-CD	-21.67	72.93	120.60
1	A	500	PHE	C-N-CA	12.24	173.41	122.00
1	A	501	PRO	CA-N-CD	-5.14	104.30	111.50

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	282	ASN	Peptide
2	P	12	DT	Sidechain

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3461	0	3544	321	0
1	B	3452	0	3531	333	0
2	E	229	0	125	15	0
2	P	273	0	147	16	0
3	F	298	0	170	11	0
3	T	336	0	192	20	0
4	A	30	0	12	1	0
4	B	30	0	12	1	0
5	A	2	0	0	0	0
5	B	2	0	0	0	0
6	A	10	0	0	4	0
6	F	1	0	0	0	0
6	T	1	0	0	0	0
All	All	8125	0	7733	696	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 44.

The worst 5 of 696 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:480:THR:HG23	1:A:483:GLU:HG2	1.22	1.14
1:B:75:THR:HG23	1:B:78:GLN:HB2	1.29	1.12
3:F:6:DG:H2"	3:F:7:DA:H5"	1.28	1.12
1:B:457:THR:HG23	1:B:471:ALA:HB2	1.34	1.08
1:B:480:THR:HG23	1:B:483:GLU:HG3	1.34	1.03

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	428/508 (84%)	312 (73%)	86 (20%)	30 (7%)	1	22
1	B	427/508 (84%)	290 (68%)	87 (20%)	50 (12%)	0	9
All	All	855/1016 (84%)	602 (70%)	173 (20%)	80 (9%)	1	14

5 of 80 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	47	SER
1	A	124	LYS
1	A	193	MET
1	A	478	VAL
1	A	501	PRO

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	384/454 (85%)	334 (87%)	50 (13%)	5	32
1	B	383/454 (84%)	324 (85%)	59 (15%)	3	24
All	All	767/908 (84%)	658 (86%)	109 (14%)	4	29

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	508	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	109	ASP
1	B	473	THR
1	B	33	LYS
1	B	70	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	340	GLN
1	A	431	GLN
1	B	379	GLN
1	A	378	GLN
1	A	379	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	8OG	F	5	3	16,25,26	1.34	2 (12%)	21,37,40	1.84	4 (19%)
3	8OG	T	5	3	16,25,26	1.29	2 (12%)	21,37,40	1.95	5 (23%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	8OG	F	5	3	-	0/3/21/22	0/3/3/3
3	8OG	T	5	3	-	0/3/21/22	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	T	5	8OG	C2-N1	2.88	1.40	1.35
3	F	5	8OG	C2-N1	2.98	1.40	1.35
3	T	5	8OG	C6-N1	3.87	1.40	1.33
3	F	5	8OG	C6-N1	4.02	1.40	1.33

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	T	5	8OG	N3-C2-N1	-5.17	119.56	127.44
3	F	5	8OG	N3-C2-N1	-5.05	119.76	127.44
3	T	5	8OG	C2'-C1'-N9	-3.69	112.12	115.83
3	F	5	8OG	C5-C6-N1	-3.21	119.20	123.59
3	F	5	8OG	C2'-C1'-N9	-3.08	112.73	115.83

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	F	5	8OG	3	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected

value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
4	DTP	A	1521	5	24,32,32	0.97	1 (4%)	32,50,50	1.15	4 (12%)
4	DTP	B	1521	5	24,32,32	0.80	0	32,50,50	1.03	3 (9%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	DTP	A	1521	5	-	0/18/34/34	0/3/3/3
4	DTP	B	1521	5	-	0/18/34/34	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1521	DTP	C8-N7	-2.17	1.30	1.34

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1521	DTP	PA-O3A-PB	-2.94	124.47	132.73
4	B	1521	DTP	PA-O3A-PB	-2.66	125.26	132.73
4	A	1521	DTP	PB-O3B-PG	-2.45	124.44	132.67
4	B	1521	DTP	PB-O3B-PG	-2.40	124.64	132.67
4	A	1521	DTP	N3-C2-N1	-2.00	127.36	128.89

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1521	DTP	1	0
4	B	1521	DTP	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	432/508 (85%)	-0.37	2 (0%) 91 86	49, 77, 117, 128	36 (8%)
1	B	431/508 (84%)	-0.14	2 (0%) 91 86	57, 106, 138, 154	62 (14%)
2	E	11/13 (84%)	-0.53	0 100 100	90, 114, 153, 153	2 (18%)
2	P	13/13 (100%)	-0.21	0 100 100	65, 103, 171, 171	0
3	F	14/18 (77%)	-0.09	0 100 100	98, 115, 169, 176	1 (7%)
3	T	16/18 (88%)	-0.27	0 100 100	80, 97, 153, 155	0
All	All	917/1078 (85%)	-0.25	4 (0%) 93 89	49, 91, 134, 176	101 (11%)

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	518	ASN	2.8
1	B	409	ARG	2.5
1	A	518	ASN	2.3
1	A	517	PRO	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	8OG	F	5	23/24	0.92	0.19	-	104,110,114,117	0
3	8OG	T	5	23/24	0.96	0.17	-	77,81,84,90	0

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
4	DTP	B	1521	30/30	0.91	0.26	0.80	135,144,156,157	0
5	CA	B	1523	1/1	0.98	0.24	0.51	94,94,94,94	0
5	CA	B	1522	1/1	0.90	0.21	0.01	88,88,88,88	0
4	DTP	A	1521	30/30	0.96	0.19	-0.41	91,100,109,110	0
5	CA	A	1522	1/1	0.95	0.15	-1.24	64,64,64,64	0
5	CA	A	1523	1/1	0.99	0.17	-1.30	67,67,67,67	0

6.5 Other polymers [i](#)

There are no such residues in this entry.